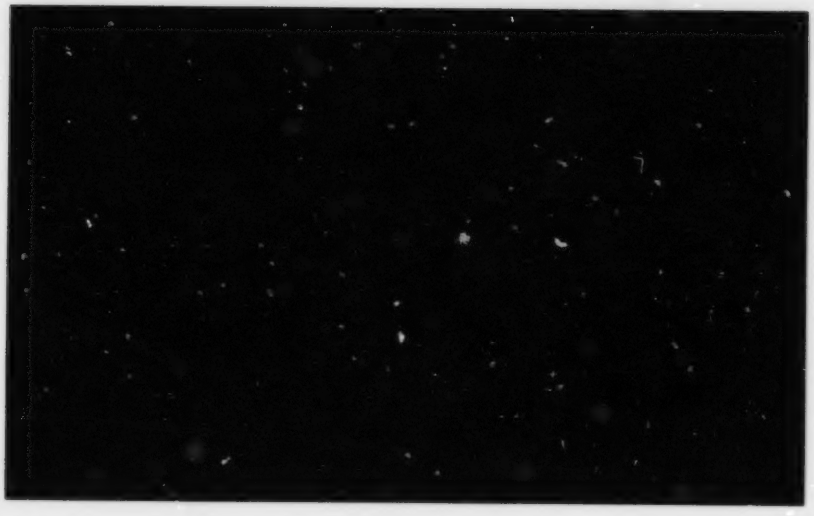
CENTRE DE RECHERCHE
SUR LES TRANSPORTS
(C.R.T.)
CENTRE FOR RESEARCH
ON TRANSPORTATION

UNIVERSITÉ DE MONTRÉAL
ÉCOLE DES HAUTES ÉTUDES COMMERCIALES
ÉCOLE POLYTECHNIQUE



CENTRE DE RECHERCHE SUR LES TRANSPORTS
UNIVERSITÉ DE MONTRÉAL
C.P. 6128, SUCCURSALE A,
MONTRÉAL
CANADA H3C 3J7

TÉLÉPHONE: (514) 343-7575
TÉLEX: (21) 05-561325
EMAIL: crt@crt.umontreal.ca
TÉLÉCOPIEUR (FAX): (514) 343-7121

91-02552/

CRT-750

A Bilevel Programming Approach to Estimating O-D Matrix by Traffic Counts

by
Michael Florian
and
Yang Chen

Centre de recherche sur les transports — Publication #750
Département d'informatique et de recherche opérationnelle — Publication #757

January 1991

Dépôt légal – 1er trimestre 1991

ABSTRACT

In this paper we present a bilevel programming formulation for the trip demand estimation (TDE) problem in a congested network where it is assumed that the flows satisfy the user optimal principle. It is possible both to provide an a priori estimate of the O-D matrix or to use only the information available from traffic counts. The most important contribution of this paper is an exact algorithm to solve the bilevel TDE problem. This algorithm uses the penalty function approach and the trust region strategy; it converges to a local minimum of the equivalent nonlinear programming problem. By using the relaxation and decomposition principles, as well as the underlying network structure, the proposed algorithm may be used to solve medium size TDE problems. Moreover, the algorithm may be applied to other convex bi-level programming problems in which the lower level problem has a network structure.

Keywords: bilevel programming, matrix estimate, equilibrium assignment, nondifferential optimization, penalty function method, trust region method.

RÉSUMÉ

Dans cet article nous présentons une formulation de programmation mathématique à deux niveaux du problème d'estimation de matrice origine-destination (EMD) pour des réseaux congestionnés où les flots satisfont le principe descriptif de choix de route. Il est possible de fournir une estimation "a priori" de la matrice de demande ou d'utiliser seulement l'information contenue dans les comptages de trafic. La contribution principale de l'article est un algorithme exact pour résoudre le problème EMD à deux niveaux. L'algorithme utilise l'approche de fonction de pénalité et la stratégie de région de confiance; il converge à un minimum local du problème équivalent de programmation nonlinéaire. En utilisant les principes de relaxation et décomposition ainsi que la structure sous-jacente de réseau, cet algorithme peut résoudre des problèmes de EMD de taille moyenne. L'algorithme peut être également appliqué à des problèmes de programmation convexes à deux niveaux quand le problème interne a une structure de réseau.

Mots-clés: programmation mathématique à deux niveaux, estimation de la matrice, affectation d'équilibre, optimisation non-différentielle, approche de fonction de pénalité, approche de région de confiance.

1 Introduction

The origin-destination (O-D) estimation (or adjustment) problem is that of "determining an estimate of the origin-destination trip demand matrix G by efficiently combining traffic count data and all other available information" (Cascetta and Nguyen (1988)). We shall refer to this problem and its many variants as the trip demand estimation problem or TDE problem.

We consider a transportation network $R = (N, A)$ with nodes $i(j) \in N$ and arcs $(i,j) \in A \subset N \times N$. The demands g_l for origin-destination pairs $l, l \in L$, induce link flows $v_{ij}, (i,j) \in A$. Traffic counts, $\hat{v}_{ij}$, are available for a subset of the links of the network $\hat{A}, \hat{A} \subset A$. Let p_{ijl} be the proportion of the trips of O-D pair l which use link $(i,j), (i,j) \in \hat{A}$. The basic equations which characterize to the problem are:

$$\sum_{l \in L} p_{ijl} g_l = \hat{v}_{ij}, \qquad (i,j) \in \hat{A}. \tag{1}$$

The assignment p_{ijl} of the matrix g_l reproduces the observed flows exactly.

The system of equations of (1) is, in general, an underspecified system. Additional information or additional criteria must be added in order to obtain a unique solution for the TDE problem. A variety of approaches has been propoed for this problem (see, for instance, Nguyen (1984)) which result in different models, depending on the available information or the additional criteria used. Nguyen (1984) surveyed the models and methods that were suggested for the TDE problem. Later Fisk (1988) classified the approaches to the problem into three groups which are:

- approaches based on the model proposed by Nguyen (1977),
- maximum entropy models based on the model of Van Zuylen and Willumsen (1980) as well as the extension to congested networks by Fisk (1987) and finally,
- models based on the combined distribution-assignment (CDA) problem of Erlander et al. (1979) and Fisk and Boyce (1983).

Brenninger-Göthe et al. (1989) proposed a multiobjective formulation of the TDE problem which takes into account the constraints related to traffic counts into the objective function.

Most of these models are formulated for uncongested networks, where travel time is independent of the flow and, since the p_{ijl} are fixed, are easier to solve. In networks that are subject to congestion, the proportion p_{ijl} are flow dependent and the TDE problem leads to model formulations which are more difficult to solve. Fisk (1987,1988) discusses the computational diffculties asociated with the TDE problem on congested networks.

In this paper we present a bilevel programming formulation of the TDE problem on a congested network, where it is assumed that the user optimal principle applies. It is the same model variety studied by Spiess (1990), who developed a gradient approach for solving the problem. Unlike other existing models, it is possible either to provide an " a priori" estimate of the O-D matrix or to use only the information available from traffic counts. The most important contribution of this paper is an exact algorithm to solve the bilevel TDE problem. This algorithm uses the penalty function approach and the trust region strategy; it converges to a local minimum of the equivlent nonlinear programming (NLP) problem. By using the relaxation and decomposition principles, as well as the underlying network structure, the proposed algorithm may be used to solve medium size TDE problems.

The paper is organized as follows. In the next section, the formulation of the TDE problem is presented and the general nature of the model is pointed out. The bilevel programming (BLP) formulation is transformed into a series of equivalent NLP problems. The last of these problems is solved by a sequence of linearly constrained penalty problems, which are described in the third section. In the fourth section, the algorithm is presented as well as a convergence proof. Some numerical results are given in the fifth section and a conclusion is given in the last section of this paper.

2 Formulation Of The User Optimal TDE Problem

Let $R = (N, A)$ be the transportation network, where N is the set of nodes and A set of links. Denote by G the set of demands of O-D pairs and $s(v)$ the cost funtion defined on link flows which is assumed to be symmetric, monotone, differentiable and seperable. Thus $s(v)$ can be expressed as

$$s(v) = (s_{ij}(v_{ij}) | (i,j) \in A), \tag{2}$$

Smith (1979) showed that the mathmatical model based on the user optimal path choice hypothesis reduces to the variational inequality (VI)

$$s(v^*)(v^* - v) \leq 0, \qquad \forall v \in \Omega(G) \tag{3}$$

where $\Omega(G)$ is the set of feasible link flow for a given demand G and $v^* \in \Omega(G)$ is the user optimal solution. Under certain conditions the VI (3) has an equivalent optimazation form which has been studied extensively,namely

$$\min f(v) = \sum_{(i,j)\in A} \int_0^{v_{ij}} s_{ij}(x)dx \tag{4}$$

$$s.t. \quad v \in \Omega(G) \tag{5}$$

The user optimal TDE problem consists in determining an estimate of the O-D demand G such that when user optimal assignment is applied, the link traffic count observations and/or prior information about the old matrix is used.

It is noted that, if the transportation network is congested, the proportions p_{ijl} are dependent on the link flows. Then, equation (1) has the form

$$\sum_{l\in L} \sum_{k\in K_l} \delta_{ijk} h_k = \hat{v}_{ij}, \qquad (i,j) \in \hat{A} \tag{6}$$

by

$$p_{ijl} g_l = \sum_{k\in K_l} \delta_{ijk} h_k, \qquad (i,j) \in \hat{A}, l \in L \tag{7}$$

where K_l is the set of paths between O-D pair l, h_k the flow on path k and δ_{ijk} is defined as

$$\delta_{ijk} = \begin{cases} 1 & \text{if path k passes through link (i,j)} \\ 0 & \text{otherwise} \end{cases} \tag{8}$$

According to the relation between path flows and link flows, equation (6) is identical to

$$v_{ij} = \hat{v}_{ij}, \quad (i,j) \in \hat{A} \tag{9}$$

which implies that there is no difference between the traffic count observation $\hat{v}$ and the values v estimated by the model. In general, it is more realistic to consider the distance between $\hat{v}$ and v as one of the objectives of the problems. In this problem, the desired trip demand matrix G is the "best"

one in the sense of the closeness between $v(G)$ and $\hat{v}$ as well as G and the "target" demand $\hat{G}$. At the same time, $v(G)$ is an optimal solution to the equilibrium assignment problem for a fixed G.

The mathematical formulation of the model as a BLP is stated as follows:

$$\min_{G \geq 0} F(G, v) = \gamma_1 F_1(G) + \gamma_2 F_2(v) \tag{10}$$

$$s.t. \quad v = assign(G) \tag{11}$$

where $F_1(G)$ is any suitable distance metric between the estimated trip demand matrix and target trip demand matrix, with some known probability distribution, and $F_2(v)$ some suitable distance metric between the observed traffic count and the desired equilibrium flow pattern. Both $F_1(G)$ and $F_2(v)$ may be taken to be quadratic functions.

In this model, γ_1 and γ_2 are predetermined weights which reflect the different beliefs in the observed traffic counts $\hat{v}$ and target demand $\hat{G}$. It is clear that if γ_2 is large enough, then $F_2(v)$ tends to zero. Therefore, the models in Van Zuylen and Willumsen (1980) and the extensions to congested case by Fisk (1987,1988) are special cases of this model if γ_1/γ_2 is small enough. Furthermore, the model contributes a new means to estimate the trip demand without making use of any "a priori" information when $\gamma_1/\gamma_2 = 0$. In the next section, a detailed description of the model is given and several equivalent NLP problems are derived.

3 Transformation From BLP Problem To NLP Problems

The BLP user optimal TDE problem can be restated as

$$\min_{G \geq 0} F(v, G) = \gamma_1 F_1(G) + \gamma_2 F_2(v) \tag{10}$$

where v solves the following problem

$$\min f(v, G) = \sum_{(i,j) \in A} \int_0^{v_{ij}} s_{ij}(x) dx \tag{12}$$

$$s.t. \qquad v \in \Omega(G) \tag{13}$$

There are more than one ways to express the feasible set $\Omega(G)$. The one used in this paper is the multicommodity flow formulation

$$\Omega(G) = \{v = \sum_{l \in L} v^l \mid \Delta^l v^l = g_l e^l, l \in L\} \tag{14}$$

where Δ^l is the node-arc incidence matrix of the network $R = (N, A)$ without the row corresponding to the destination of O-D pair l, $v^l = (v^l_{ij} | (i,j) \in A)$ is the flow vector corresponding to l and $e = (0, \ldots, 0, 1, 0, \ldots, 0)^T$ is an $|N| - 1$ dimension vector in which "1" corresponds to the origin node of l. It is clear that the row vectors of Δ^l are linearly independent for each $l \in L$.

The lower level problem (12)-(13) represents an equilibrium assignment problem with fixed G and it may be viewed as a nonlinear multicommodity flow problem. It is a convex BLP problem since the problem (12)-(13) is a convex programming problem. A general discription of convex BLP problems can be found, for instance, in Bard (1983), Bard (1988) and Bard and Falk (1982). One of the ways of solving convex BLP problems is to use the Karush-Kuhn-Tucker conditions. Therefore the Karush-Kuhn-Tucker conditions of the lower lever problem will play an important role in the process of developing a solution method for this specific convex BLP problem.

As an immediate corollary of the result in Bard (1982), we have the following elementary result:

Proposition 1 . *The BLP problem (10)-(13) is equivalent to the NLP problem defined below:*

(NLP)

$$\min F(v, G)$$

$$s.t. \quad v^l_{ij}[s_{ij}(v_{ij}) + u^l_i - u^l_j] = 0 \quad (i,j) \in A, \ l \in L \tag{15}$$

$$s_{ij}(v_{ij}) + u^l_i - u^l_j \geq 0 \quad (i,j) \in A, \ l \in L \tag{16}$$

$$\Delta^l v^l = g_l e^l \quad l \in L \tag{17}$$

$$v^l \geq 0 \quad l \in L \tag{18}$$

$$G \geq 0 \tag{19}$$

The constraints (15) and (16) of this NLP problem are the Karush-Kuhn-Tucker conditions of problem (12)-(13) where $u = (u_i^l | i \in N, l \in L)$ are the Lagrange multipliers. Note that $u_i^l = 0$ if i is the origin of O-D pair l, $l \in L$. If one can find the optimal solution to the NLP problem, then one finds the solution of the BLP problem as well, by the proposition.

However, the difficulties involved in this problem are not only the non-convexity of the induced region of the BLP problem, but also the huge number of nonlinear constraints. The idea of the surrogate constraint is adopted here to reduce the number of constrains and its application leads to the relaxed problem given below:

(RLP)

$$\min F(v, G)$$

$$s.t. \quad s_{ij}(v_{ij}) + u_i^l - u_j^l \geq 0, \quad (i,j) \in A, \quad l \in L \tag{16}$$

$$\sum_{l \in L} \sum_{(i,j) \in A} \theta_{ij}^l v_{ij}^l [s_{ij}(v_{ij}) + u_i^l - u_j^l] = 0, \tag{20}$$

$$\Delta^l v^l = g_l e^l \quad l \in L \tag{17}$$

$$v^l \geq 0 \qquad l \in L \tag{18}$$

$$G \geq 0 \tag{19}$$

Lemma 1 . *Let* $\Psi(\theta) = \min\{F(v,G) \mid (16) - (20)\}$, *then*

$$\Psi(\theta) \leq \min\{F(v,G) \mid (15) - (19)\}$$

for any $\theta \in R^{|A| \times |L|}$.

□

Lemma 2 . *Let* W *be the feasible set of the problem (10), (16)-(19) and let* $W(\theta)$ *be the feasible set of the relaxed problem. Then,*

$$W = W(\theta) \qquad \textit{for any fixed } \theta > 0.$$

Particularly,

$$W = W(\bar{1})$$
$$= \{(v, G, U) | \sum_{(i,j) \in A} v_{ij} s_{ij}(v_{ij}) + \sum_{l \in L} u_{l(i)}^l g_l = 0, \textit{ and } (16) - (19)\}$$

where $\bar{1} = (1, \ldots, 1)^T$ *and* $l(i)$ *is the destination of O-D pair* l , $l \in L$.

Proof. It is clear that if the constraints (15) are satisfied, the constraint (20) is satisfied as well. Thus $W \subseteq W(\theta)$ for any θ.

On the other hand, if (20) holds for $\theta > 0$, then it follows that

$$v_{ij}^l[s_{ij}(v_{ij}) + u_i^l - u_j^l] = 0, \qquad (i,j) \in A, \quad l \in L$$

because of constraints (16) and (18). This implies that

$$W(\theta) \subseteq W, \qquad for \quad \theta > 0,$$

and hence,

$$W(\theta) = W, \qquad for \quad \theta > 0.$$

In particular,

$$\begin{aligned}
&\sum_{l\in L}\sum_{(i,j)\in A} v_{ij}^l[s_{ij} + u_i^l - u_j^l] \\
&= \sum_{(i,j)\in A}\sum_{l\in L} v_{ij}^l s_{ij}(v_{ij}) + \sum_{l\in L}\sum_{(i,j)\in A}(u_i^l v_{ij}^l - u_j^l v_{ij}^l) \\
&= \sum_{(i,j)\in A} v_{ij} s_{ij}(v_{ij}) + \sum_{l\in L}\sum_{i\in N} u_i^l\Big(\sum_{(i,k)\in A} v_{ik}^l - \sum_{(h,i)\in A} v_{hi}^l\Big) \\
&= \sum_{(i,j)\in A} v_{ij} s_{ij}(v_{ij}) + \sum_{l\in L} u_{l(i)}^l g_l \quad .
\end{aligned}$$

Therefore,

$$\begin{aligned}
W &= W(\bar{1}) \\
&= \{(v,G,U) \mid \sum_{(i,j)\in A} v_{ij} s_{ij}(v_{ij}) + \sum_{i\in L} u_{l(i)}^l g_l = 0, \quad (16)-(19)\}
\end{aligned}$$

□

Based on the lemmas above, we have the following:

Theorem 1 . *The BLP problem (10)-(13) is equivalent to the NLP problem*

$$\begin{aligned}
&\min F(v,G) \\
&s.t. \quad s_{ij}(v_{ij}) + u_i^l - u_j^l \geq 0 \quad (i,j) \in A, \quad l \in L && (16)\\
&\qquad \Delta^l V^l = g_l e^l \qquad\qquad l \in L && (17)\\
&\qquad v_l \geq 0 \qquad\qquad l \in L && (18)\\
&\qquad G \geq 0 && (19)
\end{aligned}$$

$$\sum_{(i,j)\in A} v_{ij} s_{ij}(v_{ij}) + \sum_{l\in L} u^l_{l(i)} = 0 \qquad (21)$$

which we denote (NLPS).

The solution algorithm is designed to solve the BLP problem by solving the NLP problem (NLPS). The equivalence of (BLP) and (NLPS) motivates this approach. It is assumed that a constraint qualification holds for the problem (NLPS). This is not a strict limitation nor a strong assumption for the problem, since the matrix corresponding to the gradients of the constraints at a feasible point has the following form:

$$\begin{pmatrix}
s_{12}(v_{12}) + v_{12}s'_{12}(v_{12}), & \cdots, & s_{ij}(v_{ij}) + v_{ij}s'_{ij}(v_{ij}), & \cdots, & \vdots & 0 & \cdots, & g_l, & \cdots, & 0 \\
s'_{12}(v_{12}) & & & & & & & & & \\
 & \ddots & & & \vdots & & & & & \\
 & & s'_{ij}(v_{ij}) & & & & & (\Delta^l)^T & & \\
 & & & \ddots & \vdots & & & & & \\
\cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\
 & & \Delta^l & & \vdots & \bar{0} & & & \bar{0} & e_l \\
 & & I & & \vdots & \bar{0} & & & \bar{0} & \bar{0} \\
0 & & & & \vdots & 0 & & & 0 & 1
\end{pmatrix}$$

for all $l \in L$, where $\bar{0} = (0, \ldots, 0)^T$ and I is a suitable dimensioned identity matrix.

In practice, the row vectors corresponding to the active constraints are usually linearly independent because of the structure of Δ^l and I.

4 A Penalty Function Approach For Solving (NLPS)

In this section, we propose a penalty function method to solve the NLP problem (NLPS). The penalty problem is constructed by dualizing the non-linear constraints in the objective function as penalty terms and keeping the linear constraints. Before presenting the algorithm, some elementary relations between the local minimizers of (NLPS) and the penalty problem with linearly constraints are discussed.

The penalty problem, which we denote (PNP), is as follows:

$$\min \tilde{F}(v,G,U,\rho) = F(v,G) + \rho_0|f_0| + \sum_{l\in L}\sum_{(ij)\in A} \rho_{ij}^l (f_{ij}^l)^+ \qquad (22)$$

$$s.t. \quad \Delta^l v^l = g_l e^l, \quad l \in L \qquad (17)$$

$$v^l \geq 0, \qquad l \in L \qquad (18)$$

$$G \geq 0 \qquad (19)$$

where

$$f_0 = \sum_{(ij)\in A} v_{ij} s_{ij}(v_{ij}) + \sum_{l\in L} u_{l(i)}^l g_l \qquad (23)$$

$$f_{ij}^l = u_j^l - u_i^l - s_{ij}(v_{ij}) \qquad (i,j) \in A, \quad l \in L \qquad (24)$$

and $\rho = ((\rho_0, \rho_{ij}^l) \mid (i,j) \in A, l \in L)$ is the penalty parameter, and $(".")^+$ stands for the function $\max(0, ".")$.

Combining the results in Clarke (1976) and Hiriart-Urruty (1975), the first order necessary condition for this linearly constrained penalty problem can be stated as:

Lemma 3 . *If $(v,G,U)^*$ is a local minimizer of the penalty problem (PLP), then there exist multipliers $\lambda_k \geq 0$ such that*

$$0 \in \partial \tilde{F}((v,G,U)^*, \rho) + \sum_{k\in A^*} \lambda_k \bar{a}_k \qquad (25)$$

where $\partial \tilde{F}((v,G,U)^*, \rho)$ is the generalized gradient set of $\tilde{F}$ at $(v,G,U)^*$, A^* is the index set of the active constraints at $(v,G,U)^*$ and $\bar{a}_k'$s are the row vectors of the coefficient matrix determined by (17)-(19). □

In the case of L_1 norm exact penalty function, the generalized gradient set of $\tilde{F}((v,G,U),\rho)$ at $(v,G,U)^*$ can be determined explicitly by

Lemma 4 .

$$\partial \tilde{F}((v,G,U)^*, \rho)$$

$$= Co.\{\nabla F^*(v,G) + \rho_0 \alpha_0 \nabla f_0^*(v,G,U) + \sum_{(i,j)\in A}\sum_{l\in L} \rho_{ij}^l \alpha_{ij}^l \nabla f_{ij}^{l*}(v,G,U)\}$$

where

$$\alpha_0 \begin{cases} \in [-1,1] & \text{if } f_0^*(v,G,U)=0 \\ = sgn f_0^*(v,G,U) & \text{otherwise} \end{cases}$$

and

$$\alpha_{ij}^l \begin{cases} \in [0,1] & \text{if } f_{ij}^{l*}(v,G,U)=0 \\ = \frac{1+sgn f_{ij}^{l*}(v,G,U)}{2} & \text{otherwise} \end{cases}$$

where Co.{.} stands for the convex hull. □

We state now the fundamental result on which the algorithm is based.

Theorem 2 *Let $(v,G,U)^*$ be a feasible point of $(NLPS)$ which satisfies the first order necessary conditions of the penalty problem. Then $(v,G,U)^*$ is a KKT point of $(NLPS)$.*

Proof. The first order necessary conditions of the penalty problem are satisfied if there exist multipliers $\lambda_k \geq 0$ such that

$$0 \in \partial \tilde{F}((v,G,U),\rho) + \sum_{k \in \mathcal{A}^*} \lambda_k \bar{a}_k$$

By the definition of $\partial \tilde{F}((v,G,U),\rho)$, there exist $(\alpha_0, \alpha_{ij}^l)$ such that

$$\nabla F((v,G)^*) + \rho_0 \alpha_0 f_0^*(v,G,U) + \sum_{l \in L} \sum_{(i,j) \in A} \rho_{ij}^l \alpha_{ij}^l f_{ij}^{l*}(v,G,U) + \sum_{k \in \mathcal{A}^*} \lambda_k \bar{a}_k = 0.$$

Since $(v,G,U)^*$ is feasible to (NLPS), thus

$$f_0^*(v,G,U) = 0$$

$$f_{ij}^{l*}(v,G,U) \leq 0 \quad (i,j) \in A, l \in L.$$

By Lemma 4, $(\alpha_0, \alpha_{ij}^l)$ satisfies

$$\alpha_0 \in [-1,1]$$

$$\alpha_{ij}^l \begin{cases} \in [0,1] & \text{if } f_{ij}^{l*}(v,G,U)=0 \\ = 0 & \text{if } f_{ij}^{l*}(v,G,U)<0 \end{cases}$$

Let $\beta_0 = \rho_0 \alpha_0$, $\beta_{ij}^l = \rho_{ij}^l \alpha_{ij}^l, (i,j) \in A, l \in L$, and

$$\bar{\lambda_k} = \begin{cases} \lambda_k & \text{if } k \in \mathcal{A}^* \\ 0 & \text{otherwise} \end{cases}$$

It is clear that, for nonlinear constraints, $\beta_{ij}^l \geq 0$ and $\beta_{ij}^l f_{ij}^{l*}(v,G,U) = 0$ for $(i,j) \in A, l \in L$. For linear constraints, $\bar{\lambda}_k \geq 0, \forall k$ and $\bar{\lambda}_k = 0$ if k is in the inactive constraint set. Therefore $(\beta_0, \beta_{ij}^l, \bar{\lambda}_k)$ are the Lagrange multiplier of the problem $(NLPS)$ at $(v,G,U) = (v,G,U)^*$. This implies that $(v,G,U)^*$ is a KKT point of (NLPS). □

The basic ideas of the algorithm are, first, to use relaxation principle to deal with the penalty problem by dropping some constraints and, second, to use the trust region strategy to solve the problem. At each iteration, the Dantzig- Wolfe decomposition method is adopted to solve the resulting structured linear programming problems. The feasible point sequence generated by the algorithm converges to a KKT point of the problem (NLPS).

The linearized subproblem, which is going to be solved at each iteration, is obtained from the second order approximation of the Taylor series of (NLPS). Given the feasible point $(v,G,U)^{(m)}$ at the iteration m, the problem is defined as

$$\min l^{(m)}(W)$$

$$= F^{(m)} + \nabla^T F^{(m)} W + \rho_0 |f_0^{(m)} + \nabla^T f_0^{(m)} W| + \sum_{l \in L} \sum_{(i,j) \in A} \rho_{ij}^l [f_{ij}^{l(m)} + \nabla^T f_{ij}^{l(m)} W]^+ \quad (26)$$

$$s.t. \quad \Delta^l (X^l + v^{l(m)}) = (Y_l + g_l^{(m)}) e^l \quad l \in L \quad (27)$$

$$X^L + v^{l(m)} \geq 0 \qquad l \in L \quad (28)$$

$$Y + G^{(m)} \geq 0 \quad (29)$$

$$\| W \| \leq \xi \quad (30)$$

where

$$W = (X^{lT}, Y^T, Z^{lT}) \mid l \in L);$$

$$X^l = (x_{ij}^l \mid (i,j) \in A);$$

$$Y = (y_l \mid l \in L);$$

$$Z^l = (z_i^l \mid i \in N);$$

$$X = \sum_{l \in L} X^l;$$

and $\xi \geq 0$ is the radius of the trust region.

In the following, the objective function $F(v,G)$ is taken to be the quadratic form

$$F(v,G)=\frac{1}{2}[\sum_{(i,j)\in A}(v_{ij}-\hat{v_{ij}})^2+\sum_{l\in L}(g_l-\hat{g_l})^2]$$

The LP-like problem above can be transformed into an equivalent LP problem, $(\overline{SLP})$, as follows:

$$\min \bar{l}^{(m)}(W,p,q)$$

$$=\sum_{(i,j)\in A}(v_{ij}^{(m)}-\hat{v_{ij}})x_{ij}+\sum_{l\in L}(g_l^{(m)}-\hat{g_l})y_l+\rho_0(p_0+q_0)+\sum_{l\in L}\sum_{(i,j)\in A}\rho_{ij}^l p_{ij}^l \quad (31)$$

s.t.

$$\sum_{(i,j)\in A}(s_{ij}+v_{ij}s'_{ij})^{(m)}x_{ij}+\sum_{l\in L}u_{l(i)}^{l(m)}y_l-\sum_{l\in L}g_l^{(m)}z_{l(i)}^l-p_0-q_0=-f_0^{(m)} \quad (32)$$

$$-s_{ij}^{(m)'}x_{ij}+z_i^l-z_j^l-p_{ij}^l-q_{ij}^l=-f_{ij}^{l(m)} \quad (33)$$

$$\Delta^l X_l-y_l e^l=-\Delta^l v^{l(m)}+g_l^{(m)}e^l(=0) \quad (34)$$

$$X^l\geq -v^{l(m)} \quad (35)$$

$$\|X^l\|\leq \xi \quad (36)$$

$$Y\geq -G^{(m)} \quad (37)$$

$$\|Y\|\leq \xi \quad (38)$$

$$\|Z\|\leq \xi \quad (39)$$

$$p_0\geq 0,\ \ q_0\geq 0 \quad (40)$$

$$p_{ij}^l\geq 0,\ \ q_{ij}^l\geq 0 \quad (i,j)\in A,\ \ l\in L \quad (41)$$

where s'_{ij} stand for the derivatives of s_{ij} respectively.

It is worthwhile to point out that there are not as many constraints of (33) at each iteration as it appears, since many of the constraints (16) have been relaxed in the master problem. Consequently, there are not that many p_{ij}^l and q_{ij}^l variables in each subproblem either. The subproblem $(\overline{SLP})$ can be decomposed by O-D pairs and solved by using the Dantzig-Wolfe method. The restriction method can also be applied to deal with the variables p_{ij}^l and q_{ij}^l by setting some of them to zero. In this way the storage requirement may be reduced.

We state next the solution algorithm for solving $(\overline{SLP})$.

Solution Algorithm

Step 1.

Chose the initial feasible solution $(v, G)^{(1)}$, the penalty parameter $\rho \geq 0$, the parameters $0 < \sigma_1 < \sigma_2 < 1$ and the region radius ξ^1. Set $m := 1$.

Step 2.

The relaxed problem (RRP) of the penalty problem is:

$$\begin{aligned} \min\ & F(v, G) + \rho_0 |f_0| \\ s.t.\ & \Delta^l v^l = g_l e^l \\ & v^l \geq 0 \\ & G \geq 0 \end{aligned};$$

solve the subproblem $(\overline{SLP})$ corresponding to (RRP) at the current $(v, G, U)^{(m)}$. Denote by $(W^{(m)}, p^{(m)}, q^{(m)})$ the optimal solution to $(\overline{SLP})$.

Step 3.

Compute the actual reduction $\Delta\tilde{F}^{(m)}$ and the linearized reduction $\Delta l^{(m)}$ by

$$\begin{aligned} \Delta\tilde{F}^{(m)} &= \tilde{F}(m) - \tilde{F}((v, G, U)^{(m)} + W^{(m)}) \\ \Delta l^{(m)} &= \tilde{F}^{(m)} - l^{(m)}(W^{(m)}) \end{aligned}$$

If $\Delta l^{(m)} = 0$, go to Step 6.

Step 4.

If

$$\frac{\Delta\tilde{F}^{(m)}}{\Delta l^{(m)}} < \varphi \min(1, \frac{\Delta\tilde{F}^{(m)}}{b^{(m)}}),$$

then $\xi^{(m+1)} \in [\sigma_1 \parallel W^{(m)} \parallel, \sigma \parallel W^{(m)} \parallel]$,

else $\xi^{(m+1)} := \xi^{(m)}$. where $\varphi \in (0, \frac{1}{2}]$ is a fixed parameter and $b^{(m)}$ is a parameter determined by $\tilde{F}^{(m)}$ and $W^{(m)}$.

Step 5.

If $\Delta\tilde{F}^{(m)} \leq 0$, then

$$(v, G, U)^{(m+1)} := (v, G, U)^{(m)},$$

else $(v, G, U)^{(m+1)} := (v, G, U)^{(m)} + W^{(m)}$,

return to Step 2.

Step 6. Solve the system of inequalities

$$u_j^l - u_i^l \leq 0, \quad (i,j) \in A, \quad l \in L \tag{42}$$

$$u_i^l = \begin{cases} u_{l(i)}^{l(m)} & \text{if } i \text{ is the destination node of } l \\ 0 & \text{if } i \text{ is the origin node of } l \end{cases} \tag{43}$$

If the system has a solution, then $(v, G, U)^{(m)}$ is the desired solution. Otherwise add some constraints to the relaxed problem to update (RPP) for those arcs which do not satisfy (42). Return to Step 2.

Step 6 of the algorithm is executed only if the optimality condition of the (RRP) is satisfied. The system of inequalities (42)-(43) is not difficult to solve, since the size of the system can be reduced with the information that $u_i^l = u_{l(i)}^{l(m)}$ and $u_i^l = 0$ for origin and destination nodes respectively.

The central part of the algorithm is an extension of the algorithm of Fletcher (1989) of the linearly constrained penalty problem. The following theorems prove the convergence of the algorithm.

Theorem 3 . *If $\Delta l^{(m)} = 0$ and the system (42) – (43) has a solution, then $(v, G, U)^{(m)}$ satisfies the first order necessary conditions of the penalty problem (PNP).*

Proof. If $\Delta l^{(m)} = 0$, then

$$l^{(m)}(0) = l^{(m)}(W^{(m)}).$$

This implies that $W^{(m)} = 0$ solves the subproblem (SLP). By Lemma 3, there exist multipliers α and $\lambda \geq 0$ such that

$$\nabla F^{(m)} + \alpha_0 \rho_0 \nabla f_0^{(m)} + \sum_{l \in L} \sum_{(i,j) \in A} \alpha_{ij}^l \rho_{ij}^l \nabla f_{ij}^{l(m)} + \sum_{k \in A^*} \lambda_k \bar{a}_k = 0 \tag{44}$$

Note that the constraints $\| W \| \leq \xi^{(m)}$ are inactive constraints, since $W^{(m)} = 0$. Therefore (44) are the first order necessary conditions of (PNP), by Lemma 3.

□

Theorem 4 . *Let $\{(v,G,U)^{(m)}\}$ be the sequence generated by the algorithm when all the constraints are considered. If $\{(v,G,U)^{(m)}\}$ is bounded and $b^{(m)}$ is bounded above independently of m, then there exists a subsequence $\{(v,G,U)^{(m_k)}\}$ which converges to $(V,T,U)^{\infty}$, which satisfies the first order necessary condition of (PNP). Or equivalently, $(v,G,U)^{\infty}$ is a KKT point of $(NLPS)$.*

Proof. The proof is similar to that of Theorem 2.1 in Fletcher (1989) except

- the problem here is a constrained penalty problem and
- the relaxation procedure is used in each iteration of the algorithm.

However, if the optimality conditions for the constrained penalty problem (Lemma 3 and Theorem 6 in Hiriart-Urruty, (1975)) are used in that proof instead of considering an unconstrained problem, a similar result will be obtained. It is obvious that the relaxation procedure does not affect the convergence property since there are only a finite number of constraints (16).

□

5 A Small Example

The small example is given by one O-D pair and two links as follows:

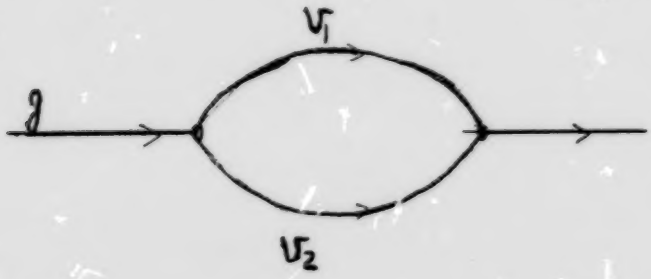


Figure 1: Small Network with Observed Flows

The cost function is defined as

$$S(v) = \begin{pmatrix} 2v_1 + 5 \\ 2v_2 + 5 \end{pmatrix}$$

The observed flow is

$$\hat{v} = \begin{pmatrix} 30 \\ 40 \end{pmatrix}$$

Suppose that the initial solution is given by $(v, G, U)^{(1)} = (20, 20, 40, 45)$. The computational result is shown in following table:

iterations	$(v, G, U)^{(m)}$	$W^{(m)}$	$\Delta\tilde{F}^{(m)}$	$\Delta l^{(m)}$	$\frac{\Delta\tilde{F}^{(m)}}{\Delta l^{(m)}}$	$\xi^{(m)}$	ρ
1	(20, 20, 40, 45)	(5, 5, 10, 10)	125	400	5/16	10	500
2	(25, 25, 50, 55)	(5, 5, 10, 10)	75	175	3/7	10	500
3	(30, 30, 60, 65)	(5, 5, 10, 10)	25	100	1/4	10	500
4	(35, 35, 70, 75)	(-0.5, -0.5, 0, 0)	-72975	25	-2919		

Table 1: Results of iterations

The optimal solution is then

$$v^* = \begin{pmatrix} 35 \\ 35 \end{pmatrix}$$

$$G^* = 70$$

6 Conclusion

In this paper, we presented a bilevel programming formulation for the TDE problem when the transportation network is congested and the route choice follows the user optimal principle. With this model it is possible to both provide an a priori estimate of the O-D matrix or to use only the information available from traffic counts. The algorithm presented above is designed to solve this complicated TDE problem. By exploiting the properties of the of the equilibrium assignment problem (the inner problem in the BLP formulation), a more compact nonlinear programming formulation permits us to develop a penalty funtion method with the trust region strategy. By

exploiting the properties of the network structure of the problem, the decomposition and relaxation principles are easily adopted so that the algoritm may be used to solve reasonable size TDE problems.

It is worthwhile to point out that the problem formulation (10)- (13) can be naturally extended to the formulation with variational inequality constraints. Consequently the algorithm can also be used to solve such problems, which are more general than the former ones in terms of the cost function. The algorithm may even be applied to general convex BLP problems which have a network structure for the lower level problem.

References

[1] Bard, J.F. (1983), An Algorithm for Solving the General Bilevel Programming Problem, *Mathematics of Operations Research* 8, 260-272.

[2] Bard, J.F. and J.E. Falk (1982), An Explicit Solution to the Multilevel Programming Problem, *Computers and Operations Research* 9, 77-100.

[3] Brenninger-Göthe, M. K.O. Jörnsten and J.T. Lundgren (1989), Estimation of Origin-Destination Matrices from Counts Using Multiobjective Programming Formulations, *Transportation Research B* 23, 257-269.

[4] Cascetta, E. and S. Nguyen (1988), A Unified Framework for Estimating or Updating Origin-Destination Matrices from Traffic Counts, *Transportation Research B* 22, 437-455

[5] Clarke, F.H. (1975), Generalized Gradients and Applications, *Trans. Amer. Math. Soc.* 205, 247-262.

[6] Clarke, F.H. (1976), A New Approach to Lagrange Multiplies, *Mathematis of Operations Research* 1, 165-174.

[7] Erlander, S., S. Nguyen and N. Stewart (1979), On the Calibration of the Combined Distribution-Assignment Model, *Transportation Research B* 13, 259-267.

[8] Fisk, C.S. (1988), On Combining Maximum Entropy Trip Matrix Estimation with User Optimal Assignment, *Transportation Research B* 22, 69-73.

[9] Fisk, C.S. (1989), Trip Matrix Estimation from Link Traffic Counts: The Congested Network Case, *Transportation Research B* 23, 331-336.

[10] Fisk, C.S. and D. Boyce (1983), A Note on Trip Matrix Estimation from Link Traffic Count Data, *Transportation Research B* 17, 245-250.

[11] Fletcher, R. (1981), *Practical Methods of Optimization*, John Wiley & Sons, New York.

[12] Fletcher, R. and E. Sainz de la Maza (1989), Nonlinear Programming and Nonsmooth Optimization by Successive Linear Programming, *Math. Programming* 43, 235-256.

[13] Hiriart-Urruity, J.B. (1975), On Optimality Conditions in Nondifferentiable Programming, *Math. Programming* 4, 73-86.

[14] Nguyen, S. (1977), Estimation an O-D Matrix from Network Data: A Network Equilibrium Approach, *Publication No. 60, CRT, Université de Montréal, Montréal, Canada*

[15] Nguyen, S. (1984), Estimating Origin-Destination Matrices from Observed Flows, In *Transportation Planning Models* (Eds. M. Florian) 363-380.

[16] Spiess, H. (1990), A Gradient Approach for the O-D Matrix Adjustment Problem, *Publication No.693, CRT, Université de Montréal, Montréal, Canada*

[17] Smith, M.J. (1979), Existence, Uniqueness and Stability of Traffic Equilibria, *Transportation Research B* 13, 295-304.

[18] Van Zuylen, J.H. and L.G. Willumsen (1980), The Most Likely Trip Matrix Estimated from Traffic Counts, *Transportation Research B* 14, 281-293.